Contribution ID: **95**   Type: **Invited oral presentation**

Valid Predictions for Diagnostic Purposes

Tuesday, 29 November 2011 11:30 (30 minutes)

The application of traditional machine-learning algorithms to modern
high-throughput and high-dimensional (with many thousands of features)
data sets often leads to serious computational diculties. Several new
algorithms, foremost support vector machines (SVM) and other kernel
methods, have been developed recently with the goal of tackling highdimensional
problems. However, a typical drawback of the new algorithms
is that they usually do not provide any useful measures of condence in
their predictions for new unclassied examples.
This talk will describe a novel technique for \hedging" predictions produced
by the new and traditional machine-learning algorithms, including,
for example, SVM, (kernel) ridge regression, (kernel) nearest neighbors,
(kernel) logistic regression, decision trees, boosting, and neural networks.
We call our algorithms for producing hedged predictions conformal predictors [1, 2].
The basic ideas of this approach are closely connected to the problem
of dening random sequences, and the ideal conformal predictor can be
constructed using the universal notion of algorithmic randomness. This
notion was dened by A. Kolmogorov, P. Martin-Lof, and L. Levin based
on the existence of universal Turing machines. Because of its universality,
algorithmic randomness is not computable, and conformal predictors
are based on computable approximations to it. It has been shown that
such approximations can be developed using typical machine-learning algorithms;
for example, for SVM the Lagrange multipliers of the support
vectors can be used to approximate the randomness level of the data set.
The talk will introduce this method, paying particular attention to the
following advantages of the hedging technique:
it gives provably valid measures of condence, in the sense that they
never overrate the accuracy and reliability of the predictions;
it does not make any additional assumptions about the data beyond
the IID assumption (the examples are independent and identically
distributed);
it allows to estimate the condence in the prediction of individual
examples;
1
conformal predictors can be used as region predictors, allowed to output
a range of labels as their prediction, so that one can control the
number of erroneous predictions by selecting a suitable condence
level;
the well-calibrated prediction regions produced by conformal predictors
can be used in both on-line and o-line modes of learning, as
well as in several intermediate modes, allowing, for example, \slow"

and \lazy" teachers.

Applications of conformal predictors to a number of real-world problems, including medical diagnosis, homeland security, automated target recognition have already been successful. The results of these applications will be presented and discussed.

References

[1] Gammerman, A., Vovk, V.: \Hedging Predictions in Machine Learning". Computer Journal, 50, No.2, 2007, pp.151-172.

[2] Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World, Springer, New York, 2005.

2

Presenter: Prof. GAMMERMAN, Alexander (University of London (Royal Holloway College))

Session Classification: Techniques of Analysis of Massive Database